PB87-145397

Estimation of Uncertainty for the 1980 NAPAP (National Acid Precipitation Assessment Program) Emissions Inventory

Alliance Technologies Corp., Bedford, MA

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

Dec 86

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing)		
1. REPORT NO. EPA/600/7-86/055	2.	3. RECIPIENT'S ACCESSION NO. PB87 145397/AS
4. TITLE AND SUBTITLE Estimation of Uncertainty for the 1980 NAPAP Emissions Inventory		5. REPORT DATE December 1986 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) John E. Langstaff and Janice K. Wagner		8. PERFORMING ORGANIZATION REPORT NO. GCA-TR-85-96-G
9. PERFORMING ORGANIZATION NAME AND ADDRESS Alliance Technologies Corporation 213 Burlington Road Bedford, Massachusetts 01730		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. 68-02-3997, Task 6
12. SPONSORING AGENCY NAME AND ADDRESS EPA, Office of Research and Development Air and Energy Engineering Research Laboratory Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED Task Final; 10/85 - 10/86 14. SPONSORING AGENCY CODE EPA/600/13

15. SUPPLEMENTARY NOTES AEERL project officer is J. David Mobley, Mail Drop 61, 919/541-2612.

16. ABSTRACT The report presents the methods, assumptions, and results of an effort to develop quantitative estimates of emissions uncertainties for the 1980 National Acid Precipitation Assessment Program (NAPAP) Emissions Inventory. (NOTE: NAPAP has charged its Task Group on Emissions and Controls with developing comprehensive and accurate inventories of natural and anthropogenic emissions of substances thought to be important in acid deposition processes. This responsibility includes quantifying the degree of uncertainty associated with those estimates.) Conventional statistical concepts were applied to sample uncertainty data. The shortcomings of these data and questions concerning the proper application of statistical methods to the problem at hand were deferred for later study. Results of this study consist of lessons learned and problems identified, and fall into three categories: uncertainty relationships at various levels of emissions disaggregation, statistical methodology questions, and gaps in input data. The computed uncertainty values are illustrative in nature and included for completeness; these values may be used only to identify uncertainty trends and relationships between the various pollutants at increasing levels of aggregation and allocation.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution Emission Inventories Precipitation (Meteorology) Acidity Probability Theory Estimating	Pollution Control Stationary Sources National Acid Precipitation Assessment Program (NAPAP)	13B 12A 14G 15E 04B 07D
18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS (This Report) Unclassified 20. SECURITY CLASS (This page) Unclassified	21. NO. OF PAGES 52 22. PRICE

EPA Form 2220-1 (9-73)

PB87-145397

EPA/600/7-86/055
December 1986

ESTIMATION OF UNCERTAINTY FOR THE 1980 NAPAP EMISSIONS INVENTORY

by

John E. Langstaff
Janice K. Wagner
ALLIANCE TECHNOLOGIES CORPORATION
(Formerly GCA Technology Division, Inc.)
213 Burlington Road
Bedford, Massachusetts 01730

EPA Contract No. 68-02-3997
Work Assignment No. 6

J. David Mobley, EPA Project Officer
Air and Energy Engineering Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

This study was conducted in cooperation with the
National Acid Precipitation Assessment Program

AIR AND ENERGY ENGINEERING RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA. 22161

NOTICE

This document has been reviewed in accordance with U. S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The National Acid Precipitation Assessment Program (NAPAP) has charged its Task Group on Emissions and Controls with developing comprehensive and accurate inventories of natural and man-made emissions of substances thought to be important in acid deposition processes. This responsibility includes quantifying the degree of uncertainty associated with those estimates. This report presents the methods, assumptions, and results of an effort to develop quantitative estimates of emissions uncertainties for the 1980 NAPAP Emissions Inventory. Conventional statistical concepts were applied to sample uncertainty data. The shortcomings of these data and questions concerning the proper application of statistical methods to the problem at hand were deferred for later study. Results of this study consist of lessons learned and problems identified, and fall into three categories: uncertainty relationships at various levels of emissions disaggregation, statistical methodology questions, and gaps in input data. The computed uncertainty values are illustrative in nature and included for completeness; these values may be used only to identify uncertainty trends and relationships among the various pollutants at increasing levels of aggregation and allocation.

CONTENTS

Abstract iii
Figure v
Tables vi
Acknowledgment vii

1. Executive Summary 1
2. Introduction 6
3. Theory, Assumptions, and Uncertainty Estimation Approach 8
   - Theory and Assumptions 8
   - Uncertainty Estimation Approach 23
4. Findings 26
   - Qualitative Uncertainty Relationships 26
   - Anomalies 26
   - Data Shortcomings 27
   - Statistical Methods 30
   - Conclusion 32

References 33

Appendix

A. 1980 NAPAP Emissions Inventory Uncertainty Workshop 34

FIGURE

Number		Page
1.	Variability in NO_x emissions as a function of boiler size	25

TABLES

Number		Page
1	Illustrative Uncertainty of the 1980 NAPAP Emissions Inventory .	3
2	Estimation of Percent Error for Emission Factors for Point Sources for Each Estimation Method	12
3	Percent Error Estimates for Area Source Emission Factors	15
4	Percent Error Estimates for Seasonal, Daily, and Hourly Allocation Factors for Point and Area Sources	17
5	Pollutant Species Allocation Carried Out in the NAPAP Inventory	19
6	Percent Error Introduced by Year-of-Record Difference for Emissions	21
7	Illustrative Uncertainty of the 1980 NAPAP Emissions Inventory .	28

ACKNOWLEDGMENT

This report describes an effort to apply a methodology to estimate the uncertainty of the 1980 NAPAP Emissions Inventory, a project administered by the U.S. Environmental Protection Agency with funding from the National Acid Precipitation Assessment Program (NAPAP) Task Group on Emissions and Controls. David J. Beecy of the U.S. Department of Energy serves as chairman of this task group, and Edward C. Trexler (DOE) serves as deputy chairman.

The contributions of several contractors to this controversial and important project are acknowledged. The uncertainty estimation methodology was developed primarily by Carmen Benkovitz and Neal Oden of Brookhaven National Laboratory. Their derivations have been detailed in extensive internal memoranda, sections of which have been included in this document. Jim Homolya of Radian documented the Uncertainty Workshop results; his report is cited extensively in this document and summarized in Appendix A. Dr. Koss Leadbetter, an EPA/AEERL statistics consultant, provided valuable statistical advice and insight. Extensive technical review assistance was provided by Steve McBrien and Brian Price of The Mitre Corporation.

SECTION 1

EXECUTIVE SUMMARY

The National Acid Precipitation Assessment Program (NAPAP) has charged its Task Group on Emissions and Controls with developing comprehensive and accurate inventories of natural and man-made emissions of substances believed to be important in acid deposition processes. The Task Group is developing estimates of past, present, and future emissions with adequate geographic, temporal, and source resolution to support the research requirements of NAPAP. The Task Group is also responsible for the quantification of the uncertainty associated with those estimates.

This report presents the methods, assumptions, and results of a study to develop quantitative estimates of emissions uncertainties for the 1980 NAPAP Emissions Inventory. Conventional statistical concepts, adapted to the emissions inventory case by Brookhaven National Laboratory in previous work, were applied to illustrative uncertainty estimates supplied by a team of emissions inventory developers. The shortcomings of these data and questions concerning the proper application of statistical methods to the problem at hand were deferred for later study.

The calculation of emissions uncertainty for individual sources and source aggregation is performed by a computer program. Annual emissions estimates for each process are allocated temporally, spatially, and by pollutant subspecies. This allocation is carried out by merging process- and pollutant-specific allocation factors to the individual emission records in the file. After these allocation factors have been assigned, the system allows the calculation of emissions by pollutant for a specified averaging time, source classification, or geographical region. The emissions uncertainty is calculated in a similar fashion. Uncertainty estimates, expressed as percent error values, are assigned to the emission records on a process- and pollutant-specific basis. By use of the error propagation formulas for products and sums, it has been assumed that emissions uncertainty can be computed directly for the various levels of source aggregation.

The initial set of uncertainty input values was provided by a panel of emissions inventory developers who were assembled at a 1980 NAPAP Emissions Inventory Uncertainty Workshop in May 1985 at the Air and Energy Engineering Research Laboratory, U.S. EPA, Research Triangle Park, North Carolina. A report detailing the results of this Workshop is included in Appendix A. At the Workshop, the panel was asked to provide consensus estimates of the

90 percent confidence interval surrounding nominal mean values for generalized classes of the variables used to calculate emissions at various levels of species, temporal, and spatial resolution.

Although members of this panel were experienced in developing emissions inventories and the methods for estimating emissions, they could not be considered experts in the calculation of emissions and associated uncertainty for all source types. They were generally not experienced with the actual emitting processes under consideration or with the typical variability of emissions from those processes.

This study was intended as a first step toward quantifying uncertainty in the 1980 NAPAP Emissions Inventory. Obvious shortcomings in the data set and doubts about validity of the methods were deferred to later study. Key assumptions in the uncertainty calculations are:

- Emission calculation parameters are independent, that is, they do not covary.
- Emission factors represent true mean values.
- All estimates are unbiased.
- The emission parameters can be treated as random variables which are approximately normally distributed.
- No coding or transcription errors are present.
- The data are complete; no emissions data, emissions sources, or emissions source categories are missing.

Results from the Workshop were later extended to other pollutants, and otherwise interpreted, by contractor personnel and the EPA Project Officer. These extensions were performed to illustrate uncertainty results for all emission estimates in the inventory. For these reasons, the results of the uncertainty calculations should be used only to identify uncertainty relationships among the various pollutants at increasing levels of aggregation and allocation.

Based upon the uncertainty estimates of the panel, the contractor, and the EPA Project Officer, and the methodology described in Section 3, illustrative emissions uncertainty values have been calculated for sulfur dioxide (SO_2), oxides of nitrogen (NO_x), speciated oxides of nitrogen ($S\text{-}NO_x$), carbon monoxide (CO), volatile organic compounds (VOC), speciated volatile organic compounds (S-VOC), total suspended particulates (TSP), speciated total suspended particulates (S-TSP), lead (Pb), ammonia (NH_3), hydrogen chloride (HCl), hydrogen fluoride (HF), and sulfate (SO_4). (See Table 1.) The percent uncertainties show an expected decrease as aggregation level increases, and an increase in uncertainty with disaggregation to finer levels of detail. It must be noted that these phenomena result inevitably from the combination of assumptions about the character of the uncertainty

TABLE 1a. ILLUSTRATIVE UNCERTAINTY OF THE 1980 NAPAP EMISSIONS INVENTORY

	Percent Error*												
	Annual												
	SO2	NOx	SNOx[1]	CO	VOC	SVOC[2]	TSP	STSP[3]	Pb	NH3	HCl	HF	SO4
Nation													
point & area	1	1	1	1	1	2	3	5	2	4	1	9	1
Region average[4]													
point & area	4	2	3	4	4	7	9	14	7	14	6	8	5
State average													
point & area	6	4	5	8	8	14	16	25	13	26	13	13	12
County average													
point & area	16	16	19	29	32	54	73	110	54	60	57	34	62
point	24	27	32	31	36	79	30	72	33	74	24	19	34
area	16	16	19	30	33	56	79	120	56	60	94	100	73
Grid average													
point & area	25	18	21	30	31	52	74	110	45	50	25	20	34
area	20	19	22	33	37	60	82	120	59	63	97	110	75
SCC average													
point	39	45	54	45	52	120	47	120	88	100	99	99	89

	Percent Error*												
	Hourly												
	SO2	NOx	SNOx	CO	VOC	SVOC	TSP	STSP	Pb	NH3	HCl	HF	SO4
Nation													
point & area	2	1	2	3	2	3	4	6	4	5	4	13	3
Region average[4]													
point & area	8	4	5	9	6	9	12	17	13	19	19	21	9
State average													
point & area	14	8	9	17	12	18	22	31	21	35	34	34	21
County average													
point & area	32	29	31	51	48	68	100	140	82	82	94	74	88
point	48	48	51	66	68	100	61	95	80	100	65	61	68
area	32	29	31	52	49	70	110	150	83	82	130	140	99
Grid average													
point & area	48	30	32	49	46	64	99	130	68	67	68	64	61
area	33	31	33	54	53	74	110	150	86	84	130	140	100
SCC average													
point	74	78	83	78	100	160	99	150	130	140	140	140	130

* With 90% confidence

1 Speciated NOx

2 Speciated VOC

3 Speciated TSP

4 EPA Region

Up to two significant digits have been reported to facilitate comparisons, and does not imply a corresponding level of precision.

TABLE 1b. ILLUSTRATIVE UNCERTAINTY OF THE 1980 NAPAP EMISSIONS INVENTORY

	Percent Error*																									
	Daily													Seasonal												
	SO2	NOx	SNOx[1]	CO	VOC	SVOC[2]	TSP	STSP[3]	Pb	NH3	HCl	HF	SO4	SO2	NOx	SNOx	CO	VOC	SVOC	TSP	STSP	Pb	NH3	HCl	HF	SO4
Nation point & area	1	1	1	2	1	2	3	5	2	4	1	10	2	1	1	1	1	1	2	3	5	2	4	1	10	1
Region average[4] point & area	5	3	3	5	4	7	9	14	8	14	9	10	6	4	2	3	4	4	7	9	14	7	14	6	8	5
State average point & area	8	5	6	9	8	14	17	26	14	27	17	18	13	6	4	5	8	8	14	16	25	13	26	14	14	12
County average point & area	21	19	22	33	34	56	77	120	58	63	63	42	65	17	16	19	30	32	55	74	110	55	61	58	35	62
point	31	33	38	37	41	83	35	75	42	78	32	28	40	25	28	33	32	36	80	31	73	35	75	25	20	35
area	21	20	22	33	36	58	82	120	59	63	98	110	77	17	16	20	31	34	56	79	120	56	61	94	100	74
Grid average point & area	32	21	23	33	33	53	77	110	49	52	34	29	38	26	18	21	30	31	52	74	110	46	50	27	22	34
area	23	22	25	36	39	61	86	130	63	65	100	110	78	20	20	22	33	37	60	82	120	59	63	98	110	75
SCC average point	50	55	62	55	61	130	56	120	94	110	110	100	95	41	47	55	47	53	120	49	120	89	100	100	100	90

* With 90% confidence

1 Speciated NOx

2 Speciated VOC

3 Speciated TSP

4 EPA Region

Up to two significant digits have been reported to facilitate comparisons, and does not imply a corresponding level of precision.

input data and the workings of the error propagation formulas on data of that character; the particular values provided by the panel are not important in this effect.

The values of uncertainty estimates for national levels of aggregation appear to be unreasonably small. The emission inventory data do not warrant this degree of confidence. Therefore, the assumptions, data, and methodology must be reexamined for reasonableness. Preliminary indications are that the methodology must be expanded to incorporate other sources of error (e.g., bias, coding, and emission errors). Other elements of the uncertainty estimates are also counterintuitive and should be corrected by improved assumptions, data, and methodology. Thus, the uncertainty estimates presented are only illustrative of the results achieved from applying the sample data set to the proposed methodology.

Despite the limitations noted above, this experiment represents a helpful step in learning how best to approach quantification of uncertainty in the NAPAP Emissions Inventories. Lessons learned here will guide further efforts toward an improved level of scientific validity and credibility. Investigation of the assumptions, methodology, and required input data is continuing.

SECTION 2

INTRODUCTION

The National Acid Precipitation Assessment Program (NAPAP) was established by Congress in 1980 to coordinate and expand research on problems posed by acid deposition in and around the United States. The program is composed of eight task groups having specific technical responsibilities. The Task Group on Emissions and Controls is charged with providing a complete and accurate inventory of anthropogenic and natural emissions thought to be important in acid deposition processes. This effort involves developing estimates of past, present, and future acid deposition precursor emissions with adequate geographic, temporal, and species resolution to support the research requirements of NAPAP. It also entails quantifying the degree of uncertainty associated with those estimates.

This report presents the methods, assumptions, and results of a study to develop quantitative estimates of emissions uncertainties for the 1980 NAPAP Emissions Inventory. In this study, conventional statistical concepts were applied to a test data set of uncertainty estimates. The results were then subjected to peer review, with the result that methodological issues, statistical questions, and data gaps have been identified and defined more precisely.

The actual calculation of emissions uncertainty for individual sources and source aggregations is performed by a computer program. Annual emissions estimates for each process are allocated temporally, spatially, and by pollutant subspecies. This allocation is carried out by merging process- and pollutant-specific allocation factors to the individual emission records in the file. After these allocation factors have been assigned, the system allows the calculation of emissions by pollutant for a specified averaging time, source classification, or geographical region. The emissions uncertainty is calculated in a similar fashion. Uncertainty estimates, expressed as percent error values, are assigned to the emission records on a process- and pollutant-specific basis. By use of the error propagation formulas for products and sums, it has been assumed that emissions uncertainty can be computed directly for the various levels of source aggregation.

Adaptation of statistical methods to the emissions inventory case was performed previously by Brookhaven National Laboratory.[1] The Brookhaven work was concerned primarily with propagation of error through a multiplicative or additive process. Emissions estimates are created by

multiplying two or more factors, for example, a unit emissions rate for a process times the number of hours the process is operated. If estimates of error can be attached to each of these factors, the corresponding uncertainty in the calculated emissions value ought to be determinable by conventional error propagation formulas. In addition, the net error in the total of all emissions from a designated class of sources could be determined by use of the error propagation formula for additive processes.

The usefulness of these methods depends heavily on the quality of data available to characterize the variability of the factors used in the emissions calculations. Statistical methods are predicated on certain data characteristics, and the error propagation formulas are no exception.

The study described here used a team of emissions inventory developers to estimate factor uncertainties. A Workshop was held at which a consensus-building procedure was employed to arrive at general uncertainty estimates for such factors as emissions rates, fuel use and economic activity, control efficiencies, and the uncertainties introduced by allocation factors. The results of this Workshop are presented in Appendix A. Contractor personnel in conjunction with the EPA Project Officer later extended the results to other pollutants by imitating patterns of relative uncertainty established by the panel.

This study therefore represents the application of a first-cut methodology to an illustrative uncertainty data set. The value of this study is not necessarily in the absolute calculated uncertainty estimates, but rather in the identification of procedural, methodological, and data issues that must be addressed in continuing work and in the investigation of uncertainty relationships among pollutants at various levels of aggregation and allocation. The lessons learned and problems identified in this study are discussed in Section 4. Section 3 describes the statistical formulations employed and the uncertainty assumptions selected by the Workshop team and contractors for use in this effort.

SECTION 3

THEORY, ASSUMPTIONS, AND UNCERTAINTY ESTIMATION APPROACH

THEORY AND ASSUMPTIONS

The statistical theory presented in this report was derived by Brookhaven National Laboratory, based on statistical approaches to uncertainty estimation and propagation of error.[1] Relevant derivations of uncertainty equations are presented below. A panel of emissions inventory developers was convened at EPA as an Emissions Inventory Uncertainty Workshop. This Workshop, held in May 1985, was designed to come to a consensus regarding the values to be used for uncertainty in the factors used to calculate total emissions. The values agreed on by the Workshop panel were used in this test of the statistical methods developed by Brookhaven since alternative data were not available.

The uncertainty of an emissions parameter is given in terms of a percent error value defined as follows. The percent error, p, of a variable X, is defined such that 90 percent of the values of X lie within the range (1-p/100)E and (1+p/100)E, where E is the mean value of X. Then if X is assumed to follow a normal distribution, the variance, V(X), of X can be given by the following equation:

$$V(X) = k^2 p^2 E^2 \qquad (1)$$

where

$$k = \frac{1}{100} \times \frac{1}{1.65},$$

and 1.65 is the 95th percentile of the standard normal distribution.[2]

Consider that X is the emission rate of a particular pollutant at the Source Classification Code (SCC) level. Assume further that the emission rate has been fully allocated with respect to time, species, and geographic area. At this fully resolved level, the emission rate corresponds to a given pollutant species emitted at the SCC level from a particular grid square at a specified hour of the day. This emission rate, X, can be expressed as the following product:

$$X = \prod_{i=1}^{N} U_i \qquad (2)$$

where

U_1 = emission factor for the point or area source SCC,

U_2 = ash or sulfur content of the fuel used, if appropriate (=1 otherwise),

U_3 = material or fuel throughput for a point source or activity level for an area source,

U_4 = penetration factor (1-percent control efficiency/100) for the pollutant (=1 for an area source),

U_5 = fraction of annual emissions which occur in the season of interest,

U_6 = fraction of seasonal emissions which occur during the day of interest,

U_7 = fraction of daily emissions which occur during the hour of interest,

U_8 = factor to convert the pollutant emission rate to the emission rate for a selected pollutant species (=1 for pollutants such as SO_2 which are not speciated),

U_9 = fraction of the emissions which fall within a given grid square (1 or 0 for a point source or a county which falls completely within or outside of a grid cell), and

$N = 9$.

Assuming that the parameters U_i are mutually independent, the variance of X can be expressed as:

$$V(X) = \prod_{i=1}^{N} \left\{ [E(U_i)]^2 + V(U_i) \right\} - \prod_{i=1}^{N} [E(U_i)]^2 \qquad (3)$$

where

$E(U_i)$ = mean value of U_i, and

$V(U_i)$ = variance of U_i.

Equation (3) is derived as follows:

$$V(X) = EX^2 - (EX)^2$$

$$= E\left(\prod_{i=1}^{N} U_i\right)^2 - \left(E\prod_{i=1}^{N} U_i\right)^2$$

$$= \prod_{i=1}^{N} EU_i^2 - \left(\prod_{i=1}^{N} EU_i\right)^2 \qquad \text{since } U_i \text{ are independent}$$

$$= \prod_{i=1}^{N} \left[(EU_i)^2 + V(U_i)\right] - \left(\prod_{i=1}^{N} EU_i\right)^2 \qquad \text{since } V(U_i) = EU_i^2 - (EU_i)^2.$$

It follows immediately from Equations (1) and (3) and the fact that

$$E(X) = \prod_{i=1}^{N} E(U_i) \tag{4}$$

that the percent error for X can be written in terms of the percent errors for the U_i's as:

$$p^2 = \frac{1}{k^2}\left(\prod_{i=1}^{N} (1+k^2 p_i^2) - 1\right) \tag{5}$$

where

p = percent error for X, and

p_i = percent error for U_i.

In the discussion which follows, each of the terms in Equation (2) will be defined more precisely and the nature of its uncertainty examined. It is important to note that this uncertainty analysis makes the following assumptions:

- Emission calculation parameters are independent, that is, they do not covary.
- Emission factors represent true mean values.
- All estimates are unbiased.

- Emission parameters can be treated as random variables which are approximately normally distributed.
- No coding or transcription errors are present.
- The data are complete; no emissions data, emissions sources, or emissions source categories are missing.

These errors, while important, will be the subject of future investigations.

The Workshop panel mentioned above came to decisions on the uncertainty for each of the factors included in Equation (2) for the following three pollutants: sulfur dioxide (SO_2), oxides of nitrogen (NO_x), and volatile organic compounds (VOC). Subsequently, the contractor and EPA Project Officer extended the analysis to include the following additional pollutants: total suspended particulates (TSP), carbon monoxide (CO), lead (Pb), ammonia (NH_3), hydrogen chloride (HCl), hyrogen fluoride (HF), and sulfate (SO_4). Although members of this panel were experienced in the development of emission inventories, they could not be considered experts in the calculation of emissions and associated uncertainty for all source types. For this reason, the results of uncertainty analyses, based upon the error estimates developed by the panel, must be regarded as an illustrative application of results of the statistical methods proposed by Brookhaven.

The remainder of this section illustrates the process used to obtain the uncertainty values which were used in this study, along with a discussion of the propagation of percent error associated with individual sources as those sources are aggregated and allocated.

Emission Factor (U_1)

The most conventional technique for the estimation of uncontrolled emissions from a given process is the use of an (AP-42) emission factor.[3] In the case of a nonfuel-burning process, the annual emissions are computed as the product of an emission factor (U_1) and the throughput for the process (U_3). For fuel-burning sources the ash or sulfur content (U_2) of the fuel is sometimes required, depending upon the particular SCC. Within EPA, use of the AP-42 emission factors to estimate emissions from a specific source category is referred to as Estimation Method 3. For Estimation Method 3 the Workshop panel based uncertainty estimates for emissions factors on the AP-42 ratings of the quality of the tests on which the AP-42 emission factors are based. It should be noted that emission factors do not reflect the variability of actual emissions within a given source category.

Nevertheless, lacking better information on the uncertainty associated with the AP-42 emission factors, the Workshop panel assigned uncertainty estimates for those emission factors on the basis of the AP-42 subjective indicators of the quality of test results. This work was extended by the contractor and the EPA Project Officer for those pollutants not considered by the panel; many of these pollutants are not included in AP-42. The results of the panel's deliberations and the extension work are presented in Table 2.

TABLE 2. ESTIMATION OF PERCENT ERROR FOR EMISSION FACTORS FOR POINT SOURCES FOR EACH ESTIMATION METHOD

Pollutant	Estimation method	Percent error				
		10	25	50	75	100
SO_2	1. Stack Test		X			
	2. Material Balance	X				
	3. AP-42 Emission Factor Rating					
	A(1)	X				
	B(2)		X			
	C(3)			X		
	D(4)				X	
	E(5)					X
	5. State		X			
	4,6,7,0 and blank					X
NO_x	1. Stack Test		X			
	2. Material Balance		X			
	3. AP-42 Emission Factor Rating					
	A(1)	X				
	B(2)		X			
	C(3)			X		
	D(4)				X	
	E(5)					X
	5. State			X		
	4,6,7,0 and blank					X
VOC	1. Stack Test		X			
	2. Material Balance			X		
	3. AP-42 Emission Factor Rating					
	A(1)	X				
	B(2)		X			
	C(3)			X		
	D(4)				X	
	E(5)					X
	5. State			X		
	4,6,7,0 and blank					X
NH_3*	All Methods					X
Other*	1. Stack Test		X			
	2. Material Balance		X			
	3. AP-42 Emission Factor Rating					
	A(1)	X				
	B(2)		X			
	C(3)			X		
	D(4)				X	
	E(5)					X
	5. State			X		
	4,6,7,0 and blank					X

Other species include CO, TSP, Pb, HCl, HF and SO_4

*Percent error values for these pollutants were developed by the contractor and EPA Project Officer.

The remaining estimation methods, 1, 2, 4, 5, 6, 7, 0, and blank, listed in Table 2 pertain to those cases in which emissions have been estimated at the point level, in spite of the fact that several SCCs could be contributing to the total emissions of a pollutant from that point. Estimation method 4 indicates a guess or unsubstantiated estimate was made. Method 6 refers to facilities under construction but not yet operational; 7 refers to facilities closed, operations ceased. Method 0 applies when emissions are known to be zero; blank occurs when the method is not reported. The percent error values, listed in Table 2, for these estimation methods pertain to the total pollutant emissions through that point. The uncertainty estimates for emissions derived from each of these estimation methods were determined by the Workshop panel based on intuitive professional judgment. Thus, the panel felt that emissions estimates based on stack tests were substantially more certain than those based on other methods. Although emission factors may be used in the calculation of point level emissions for Estimation Methods 1, 2, 4, 5, 6, 7, and 0, the quoted percent error values for these methods represent errors in the emission estimates themselves, not errors for emission factors. Since emission data are dealt with at the SCC level, these point level percent error estimates must be allocated to the SCC level. If Estimation Method 3 is not chosen, the estimated point emissions are allocated to the SCC level by use of the emission factor, throughput, and the fuel sulfur/ash content coded on the SCC record. It is then necessary to account for the uncertainty caused by this allocation process. By use of Equation (5), the percent error for the point source emission allocated to the SCC level can be written as follows:

$$Pa_j^2 = \frac{1}{k^2}\left[\left(1+k^2P_p^2\right)\left(1+k^2P_j^2\right) - 1\right] \qquad (6)$$

where

Pa_j = percent error for emissions allocated to SCC j,

P_p = percent error in point source emissions, and

P_j = percent error in the factor used to allocate the point source emissions to SCC j.

If one assumes that the true mean for the factor used to allocate the point source emissions to SCC j is given by the ratio (f_j) of the allocated emissions (A_j) to the total point emissions $\left(\sum_{j=1}^{N} A_j\right)$, then the upper bound for P_j is given by

$$P_j \leq \frac{1}{k}\sqrt{\frac{1-f_j}{f_j}} \qquad (7)$$

This inequality can be derived by considering the variance of an unbiased estimate $\hat{f}_j$ of f_j. If $0 \leq \hat{f}_j \leq 1$, then $V(\hat{f}_j)$ is maximized when $\hat{f}_j = 1$ for a proportion θ of the time, and $\hat{f}_j = 0$ for the remaining $1-\theta$ of the time. Then $E(\hat{f}_j) = \theta$ and $V(\hat{f}) = \theta(1-\theta)$. Since $\hat{f}_j$ is unbiased, $\theta = f_j$, and we have as the maximum possible variance $V(\hat{f}_j) = f_j(1-f_j)$. By Equation (1), $p_j = \frac{V(\hat{f}_j)}{k^2 f_j^2}$,

hence $p_j^2 \leq \frac{f_j(1-f_j)}{k^2 f_j^2}$, and inequality (8) follows.

In practice, we do not have the true fractions f_j, but have estimates of f_j, which we use in the inequality (7). This yields an approximate upper bound for p_j, which can either overestimate or underestimate the upper bound given by the inequality (7).

Upon substituting this upper bound in Equation (6), the following estimate for Pa_j is obtained:

$$Pa_j = \frac{1}{k}\sqrt{\left(1 + k^2 P_p^2\right)\left(1 + \frac{1-f_j}{f_j}\right) - 1} \qquad (8)$$

In the case of county area source emissions, U_1 becomes the area source emission factor for a particular area source category (e.g., Onsite Incineration--Residential). Each of those categories has been assigned an SCC to facilitate the processing of the area source data. County area source emissions are calculated by multiplying an emission factor by the associated level of activity, which is represented by the parameter U_3 in Equation (2). The percent error in these area source emission factors has been estimated separately for mobile and other area source categories. These estimates were developed by the Workshop panel, the contractor, and the Project Officer in a fashion similar to that for the point source estimates. Personnel at EPA's Office of Air and Radiation in Ann Arbor, Michigan have recommended that the following percent error estimates be used for mobile source emission factors:

VOC	50 percent
CO	29 percent
NO_x	31 percent

The area source percent error estimates are presented in Table 3.

Fuel Ash or Sulfur Content (U_2)

If an emission factor U_1 is to be used in the emission rate calculation at the SCC level and the process involves the burning of fuel, then the fuel ash or sulfur content is sometimes required for the calculation of TSP and

TABLE 3. PERCENT ERROR ESTIMATES FOR AREA SOURCE EMISSION FACTORS

	Percent error				
	10	25	50	75	100
Mobile Sources					
SO_2		X			
NO_x		X[a]			
VOC			X[b]		
*TSP					X
*CO		X[c]			
*Pb				X	
*NH_3					X
*HCl					X
*HF					X
*SO_4					X
"Other" Sources					
SO_2		X			
NO_x		X			
VOC					X
*TSP					X
*CO					X
*Pb					X
*NH_3					X
*HCl					X
*HF					X
*SO_4					X

*Percent error values for these pollutants were developed by contractor personnel and the EPA Project Officer.

[a]Office of Air and Radiation recommended 31 percent.

[b]Office of Air and Radiation recommended 50 percent.

[c]Office of Air and Radiation recommended 29 percent.

SO_2 annual emission rates. Lacking more detailed information, the panel has recommended the use of 10 percent as the percent error in the estimation of both fuel sulfur content and ash content.

Throughput or Activity (U_3)

If Estimation Method 3 has been selected for the calculation of point source emissions on the SCC level, then the material or fuel throughput (U_3) must be specified. Lacking more detailed information, the Workshop panel has assigned a percent error of 15 percent for this throughput parameter for all pollutants. At present, this percent error estimate applies to all SCCs. For county area source emissions, U_3 is the activity level. This parameter has been assigned a 25 percent error for all pollutants for both mobile and nonmobile source categories.

Penetration Factor (1 - Percent Control Efficiency/100) (U_4)

The Workshop panel recommended a value for uncertainty in control equipment removal efficiency of 25 percent for all pollutants for all SCCs. The percent error, p_c, for the penetration factor is calculated from the percent error in control efficiency as follows:

$$p_c' = \frac{p_c C}{1-C} \tag{9}$$

where

C = control efficiency divided by 100,

p_c = percent error in the control efficiency (currently assumed to be 25 percent), and

p_c' = percent error for 1-C.

Temporal Allocation Factors

For the 1980 NAPAP Emissions Inventory it has been necessary to establish temporal allocation factors to disaggregate annual values to seasonal, daily and hourly values. For most source categories (other than electric utilities) there is little data available to either support this disaggregation or to estimate the uncertainty associated with the disaggregation process. The Workshop panel recommended arbitrary uncertainty estimates for each of the disaggregation factors, taking into account the information available for utilities.

Annual to Season Emission Allocation Factor (U_5)

The fraction of annual emissions which should be apportioned to a given season has been specified as a function of point and area source SCC. Percent error estimates developed by the panel for point and area source allocation factors are given by pollutant in Table 4.

TABLE 4. PERCENT ERROR ESTIMATES FOR SEASONAL, DAILY, AND HOURLY ALLOCATION FACTORS FOR POINT AND AREA SOURCES

	Percent error				
	10	25	50	75	100
Seasonal factors					
SO_2	X				
NO_x	X				
VOC	X				
*TSP	X				
*CO	X				
*Pb	X				
*NH_3	X				
*HCl	X				
*HF	X				
*SO_4	X				
Daily factors					
SO_2		X			
NO_x		X			
VOC		X			
*TSP		X			
*CO		X			
*Pb		X			
*NH_3		X			
*HCl		X			
*HF		X			
*SO_4		X			
Hourly factors					
SO_2			X		
NO_x			X		
VOC				X	
*TSP				X	
*CO				X	
*Pb				X	
*NH_3				X	
*HCl				X	
*HF				X	
*SO_4				X	

*Percent error values for these pollutants were developed by contractor personnel and the EPA Project Officer.

Season to Day Emission Allocation Factor (U_6)

For each season, factors for allocation of season emissions to each of the three types of days of the week (Saturday, Sunday, and weekday) have been developed as a function of SCC. Percent error estimates for point and area source daily allocation factors are given in Table 4.

Day to Hour Emission Allocation Factor (U_7)

For the three types of days of the week (Saturday, Sunday, and weekday), factors have been developed to apportion daily emissions to hour of the day (given in terms of local time). Percent error estimates for hourly allocation factors are given in Table 4.

Pollutant Species Allocation Factor (U_8)

Based upon the SCC for a point or area source, TSP, NO_x, and VOC emissions are allocated to individual pollutant species as shown in Table 5. Lacking more detailed information, the Workshop panel recommended the use of a percent error of 25 percent for NO_x speciation factors and 100 percent for both VOC and TSP speciation.

Spatial Allocation Factor (U_9)

Point source emissions are allocated to grid cells (1/6 degree latitude by 1/4 degree longitude) based upon the reported location of the source. The allocation of county area source emissions to these grid cells is more complicated.

Area source emissions which are associated with the surrogate variables of population and housing are allocated to grid cells by use of population and housing data from EPA's STF3A census tapes. These tapes break down county population and housing totals to subcounty levels (township, place, tract, block, or election district). Since these subcounty entities are identified by a specific location in terms of latitude-longitude, the population and housing associated with each subcounty level is assigned to a particular grid cell. The fraction of a county's population contained within a given grid square is then calculated and used as the spatial allocation factor for that county's area source emissions associated with population. The same method is used to calculate county area source allocation factors for emissions related to housing.

The county to grid allocation factors for emissions related to land use are calculated by use of the following relationship:

$$SPAFCT_{CSi} = \frac{(A_{Ci})(A_{Si})}{\sum_{i}^{N} (A_{Ci})(A_{Si})} \qquad (10)$$

TABLE 5. POLLUTANT SPECIES ALLOCATION CARRIED OUT IN THE NAPAP INVENTORY

Pollutant	Allocated species
TSP	• Ca, Mg, Na, K, all other classes
NO_x	• NO, NO_2
THC	• Methane • Ethane • Propane • N-Butane • Isobutane • 2,3-Dimethylbutane • Pentane • Isopentane • Other Alkanes • Ethylene • Propylene • Isobutene • Trans-2 Butene • Other Alkenes • Benzene • Toluene • Xylene • Ethyl Benzene • Other Aromatics • Formaldehyde • Acetaldehyde • Propionaldehyde • Other Aldehydes • Acetone • Other Ketones • Formic Acid • Acetic Acid • Other Organic Acids

where

$SPAFCT_{CSi}$ = the spatial allocation factor for county C, land use type S, and grid i,

A_{Ci} = the portion of county C that falls within grid i,

A_{Si} = the portion of grid i with land use type S, and

N = the total number of grids covering county C.

The values for A_{Si} were obtained through an analysis of Landsat mosaic images. The AC_i parameters were calculated by use of a mapping program and an input file containing digitized county boundaries.

Lacking more detailed information, the Workshop panel has recommended the use of a 25 percent error for all spatial allocation factors, regardless of SCC. This percent error estimate is also meant to include the error due to the fact that a surrogate (such as urban land) may not be a suitable indicator of area source emissions from a source category (such as onsite industrial incineration). Also, the use of a universal percent error does not account for a case in which a large fraction of a county's area actually falls within a single grid or the case in which only one of the grids touching the county has the particular land use in question.

Year of Record Uncertainty

If an inventory is to be developed for a particular year, say 1980, one must deal with the practical difficulty that the year of record for the data for a given facility may not be 1980, but some earlier year (e.g., 1976). This would not matter if the emission characteristics of the facility have not changed. It is more likely, however, that emission characteristics have changed, but these changes have not been reported. In some cases the plant may even have ceased operation. If only data from a more recent year (e.g., 1983) are available, then these data would represent a change in at least some of the emission parameters from their 1980 values. The plant may even have come on line since 1980. On the other hand, the 1983 data may still be appropriate for 1980, but have been resubmitted by the state agency simply as a matter of course.

The calculated uncertainty is adjusted for non-1980 records by the use of an additive year of record error ε, which has zero mean since we are assuming that no bias is introduced by using non-1980 years of record. The values recommended by the Workshop panel for the uncertainty associated with ε are given in Table 6 as percent errors of the 1980 emission rate. These values reflect a belief that emissions characteristics for years near 1980 are more likely to resemble true 1980 characteristics than values for years further away.

TABLE 6. PERCENT ERROR INTRODUCED BY YEAR-OF-RECORD DIFFERENCE FOR EMISSIONS*

	Year of record difference uncertainty				
Year of record	10	25	50	75	100
1969					X
1970					X
1971				X	
1972				X	
1973			X		
1974			X		
1975		X			
1976		X			
1977		X			
1978	X				
1979	X				
1981	X				
1982	X				
1983		X			
1984		X			

*Inventory year is 1980.

In the discussion which follows, the year 1976 has been used to represent any non-1980 year of record. Assume that the percent error, p, calculated by Equation (5) represents the percent error associated with the estimate of an SCC level emission rate for the year 1976. If q is the percent error introduced by the year of record difference from 1980, and r is the percent error associated with the use of the estimated 1976 emission rate as an estimate of the 1980 emission rate, then the following approximate relationship holds:

$$r \approx \sqrt{p^2 + q^2} \tag{11}$$

To derive Equation (11), note that from Equation (1) and the above definitions of p, q and r, we have

$$V(X_{76}) = k^2 p^2 E_{76}^2,$$

$$V(\varepsilon) = k^2 q^2 E_{80}^2,$$

$$V(X_{76} + \varepsilon) = k^2 r^2 E_{80}^2,$$

where k is defined as in Equation (1),

X_{76} is the estimated emisson rate for 1976, and

E_{76}, E_{80} are the actual emission rates for 1976 and 1980, respectively.

Now, assuming that X_{76} and ε are independent,

$$V(X_{76} + \varepsilon) = V(X_{76}) + V(\varepsilon).$$

Thus, $k^2 r^2 E_{80}^2 = k^2 p^2 E_{76}^2 + k^2 q^2 E_{80}^2$.

Therefore

$$r^2 = p^2 \frac{E_{76}^2}{E_{80}^2} + q^2, \tag{12}$$

and $r^2 \approx p^2 + q^2$ when $E_{76} \approx E_{80}$.

Note that if emissions have increased from 1976 to 1980, $E_{80} > E_{76}$, then r is overestimated by Equation (11), and conversely.

UNCERTAINTY ESTIMATION APPROACH

The methodology for the assignment of percent error values for SCC level emissions discussed above has been implemented in a computer program which allows emissions uncertainty to be calculated as a function of source type, geographical location, time, and pollutant species.

Before presenting the uncertainty analysis results, it is useful to examine the properties of the percent error of a sum. Let X_i be a variable with a mean value of E_i and a percent error of p_i. The percent error, p_i, is defined such that 90 percent of the values of X_i lie within the range $(1-p_i/100)E_i$ and $(1+p_i/100)E_i$, where X_i is assumed to follow a normal distribution. With this definition of percent error, the variance, $V(X_i)$, of X_i can be given by the following equation:

$$V(X_i) = k^2 p_i^2 E_i^2 \tag{13}$$

where

$$k = \frac{1}{165}.$$

If X_T is the sum of X_i, then the total variance, V_T, is given by

$$V_T = \sum_{i=1}^{N} V(X_i) \tag{14}$$

Using the percent error definition given by Equation (13), the variance formula given by Equation (14) becomes

$$p_T^2 E_T^2 = \sum_{i=1}^{N} p_i^2 E_i^2 \tag{15}$$

where

$$E_T = \sum_{i=1}^{N} E_i$$

p_T = percent error of the total

In the calculation of emissions uncertainty, these summations are carried out over broad categories (SCC, geographical area, etc.) using SCC level emission records. For purposes of illustration, assume that each of the individual percent error values, p_i, are equal to the same value p. (In the calculation of the percent error values for the emissions totals, the actual individual percent error values for each SCC are used.) With this assumption, Equation (15) may then be rewritten as follows:

$$P_T^2 = p^2 \frac{\sum_{i=1}^{N} E_i^2}{E_T^2} \qquad (16)$$

If a large fraction of the total emissions in a category is due to a single source, then the percent error for the total will be approximately equal to the percent error for the single source (i.e., $P_T = p$). If, on the other hand, each member contributes an equal amount to the emissions sum, the percent error of the total reaches a minimum value of $\frac{p}{\sqrt{N}}$. In general, the percent error for a sum with a large number of roughly equal components will be considerably less than the percent error associated with a given member of the sum. One should keep in mind, however, that this "averaging out" process pertains only to errors which are random in nature.

Note that a common emission factor is applied for all points within a given SCC for a given species. As a result of this, the calculated point level emissions may not be independent of each other. For simplicity, we have assumed that these emissions and their associated errors are independent.

The grid-average percent errors for combined point and area sources were calculated using a further approximation, and therefore are not strictly comparable to the remaining values in Tables 1 and 7. These were approximated by

$$P_{pa} \approx \sqrt{P_p^2 E_p^2 + P_a^2 E_a^2} \Big/ (E_p + E_a), \qquad (17)$$

where

P_{pa} is the approximate grid-average percent error,

P_p is the grid-average percent error for point sources,

E_p is the total emissions from point sources,

P_a is the grid-average percent error for area sources, and

E_a is the total emissions from area sources.

The approach used in this study does not account for several factors, including the variability in emissions due to operating conditions. This can be illustrated by the data presented in Figure 1, which shows the change in flue gas NO content as a function of coal fired boiler size at a fixed excess oxygen level of 3 percent.[4] The oxygen content and other operating conditions also have a significant effect on NO_x emissions. The AP-42 emission factors do not account for these variations in NO_x emissions.

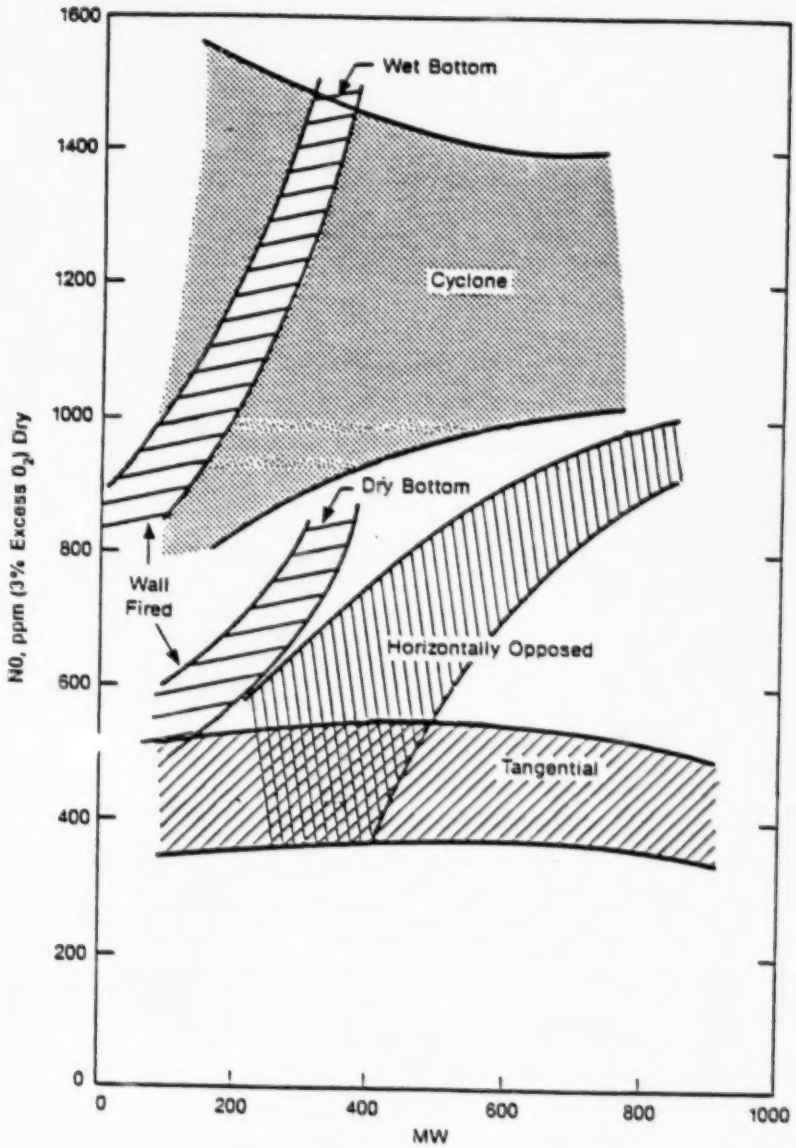

Figure 1. Variability in NO_x emissions as a function of boiler size.[4]

TABLE A-1 (continued)

J. Estimation of uncertainties of spatial resolution factors for area sources of SO_2, NO_x, and VOC.

	Uncertainty Range				
	(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
SO_2		X			
NO_x		X			
VOC		X			

SECTION 4

FINDINGS

This study of the method for estimating uncertainty associated with an emissions inventory uncovered several lessons concerning the existing capability to conduct such an analysis. These lessons are focused in three areas: qualitative relationships among the uncertainties of pollutants at varying levels of aggregation and allocation, anomalies, data shortcomings, and difficulties with the application of statistical theories in uncertainty estimates of this type. Each of these areas is discussed below.

QUALITATIVE UNCERTAINTY RELATIONSHIPS

Despite the fact that the initial uncertainty estimates are based on factors which were developed not from test data but from the experience of the Workshop panel members, the contractor and the EPA Project Officer, certain valuable information can be gained from the results that the application of the methodology produces. These results are presented in Table 7. This study implies that, overall, SO_2 and NO_x emissions estimates are less uncertain than VOC and TSP, and that ammonia and HF emissions are the most uncertain of all unspeciated pollutants.

Speciation of VOC and TSP introduces substantial uncertainty to the emissions estimates, especially at the point level. Spatial allocation of national emissions to smaller geographical areas results in increasing uncertainty estimates, an expected result of the uncertainty methodology. Similarly, as one temporally allocates emissions to shorter time periods, uncertainty increases.

ANOMALIES

In the calculations performed in this test of the uncertainty estimation methodology, a series of anomalies occurred. For the most part, these anomalies appeared as counter-intuitive results. These apparent anomalies are shown to result from the statistical methodology employed and of the initial estimates of the uncertainties of the emissions parameters.

Such anomalies include the low estimate of uncertainty at the national level. This is thought to result primarily from the assumption of a zero-mean random distribution of the errors. By the time all errors -- both positive and negative -- have been assimilated over several thousand sources, random errors tend to cancel each other out and become very small. Methodology to include omissions, coding errors, bias, and dependence of variables will all increase the national level uncertainty and could correct this problem in the future.

Another anomaly is the relatively high uncertainty of HF at the National level. This is due to a large fraction of the HF emissions coming from a few counties. The uncertainties at these counties dominate the National uncertainty and there is less cancellation of random errors. The extent to which a few counties influence the National uncertainty can be measured by the ratio of p_{T2}^2 to p_T^2, where p_T^2 is the squared aggregated percent error given by Equation (15) with the summation taken over all 3,080 counties, and p_{T2}^2 is the contribution to p_T^2 by the two counties with the highest emissions, E_1 and E_2.

$$p_{T2}^2 = \sum_{i=1}^{2} p_i^2 E_i^2 \Big/ E_T^2,$$

where p_1 and p_2 are the percent errors of the corresponding counties, and E_T is the National total emissions. For HF it turns out that the percent contribution to the squared National percent error by two counties is 95 percent.

Some apparent anomalies result from the averaging of percent error, carried out for all entries in Table 7 except at the National level. For example, at the county level, the values in Table 7 are straight averages of the county percent errors. This gives an indication of the general level of error at the county level. However, the percent error at each county is given equal weight, without regard to the contribution of each county to the total emission rate. This results in, e.g., the average percent error at the county level of SO_2 from point sources (24 percent) being larger than the average percent error of SO_2 from area sources (16 percent), which is clearly counterintuitive. If one looks at averages of county percent errors weighted by county emission totals, then this anomaly disappears: the weighted county average SO_2 percent error from point sources is 13 percent and from area sources is 19 percent.

When aggregating from the grid to the county levels, the percent errors for combined point and area sources occasionally increase. This anomaly is due to the use of the approximation Equation (17). Another apparent anomaly occurs when aggregating annual SO_2 and NO_x emissions from the county level to the state level; the SO_2 percent error becomes unexpectedly larger than the NO_x percent error.

These and other anomalies need to be addressed further before the results of this method of estimating uncertainty can be used effectively by decision-makers. Furthermore, it seems likely that the methodology needs to be modified in some manner to account for the fact that the addition of terms does not necessarily add uncertainty, and to allow for a better integration of the information concerning uncertainties from individual source categories.

DATA SHORTCOMINGS

Additional data are required to effectively use this uncertainty estimation methodology method. To some extent, additional data is required for each of the factors identified in Section 3 as contributing to both the initial emissions estimate and to the uncertainty of that estimate. These data needs include:

TABLE 7a. ILLUSTRATIVE UNCERTAINTY OF THE 1980 NAPAP EMISSIONS INVENTORY

	Percent Error*												
	Annual												
	SO2	NOx	SNOx[1]	CO	VOC	SVOC[2]	TSP	STSP[3]	Pb	NH3	HCl	HF	SO4
Nation point & area	1	1	1	1	1	2	3	5	2	4	1	9	1
Region average[4] point & area	4	2	3	4	4	7	9	14	7	14	6	8	5
State average point & area	6	4	5	8	8	14	16	25	13	26	13	13	12
County average point & area	16	16	19	29	32	54	73	110	54	60	57	34	62
point	24	27	32	31	36	79	30	72	33	74	24	19	34
area	16	16	19	30	33	56	79	120	56	60	94	100	73
Grid average point & area	25	18	21	30	31	52	74	110	45	50	25	20	34
area	20	19	22	33	37	60	82	120	59	63	97	110	75
SCC average point	39	45	54	45	52	120	47	120	88	100	99	99	89

	Percent Error*												
	Hourly												
	SO2	NOx	SNOx	CO	VOC	SVOC	TSP	STSP	Pb	NH3	HCl	HF	SO4
Nation point & area	2	1	2	3	2	3	4	6	4	5	4	13	3
Region average[4] point & area	8	4	5	9	6	9	12	17	13	19	19	21	9
State average point & area	14	8	9	17	12	18	22	31	21	35	34	34	21
County average point & area	32	29	31	51	48	68	100	140	82	82	94	74	88
point	48	48	51	66	68	100	61	95	80	100	65	61	68
area	32	29	31	52	49	70	110	150	83	82	130	140	99
Grid average point & area	48	30	32	49	46	64	99	130	68	67	68	64	61
area	33	31	33	54	53	74	110	150	86	84	130	140	100
SCC average point	74	78	83	78	100	160	99	150	130	140	140	140	130

* With 90% confidence

1 Speciated NOx

2 Speciated VOC

3 Speciated TSP

4 EPA Region

Up to two significant digits have been reported to facilitate comparisons, and does not imply a corresponding level of precision.

TABLE 7b. ILLUSTRATIVE UNCERTAINTY OF THE 1980 NAPAP EMISSIONS INVENTORY

	Percent Error*																							
	Daily													Seasonal										
	SO2	NOx	SNOx[1]	CO	VOC	SVOC[2]	TSP	STSP[3]	Pb	NH3	HCl	HF	SO4	SO2	NOx	SNOx	CO	VOC	SVOC	TSP	STSP	Pb	NH3	HCl	HF	SO4
Nation point & area	1	1	1	2	1	2	3	5	2	4	1	10	2	1	1	1	1	1	2	3	5	2	4	1	10	1
Region average[4] point & area	5	3	3	5	4	7	9	14	8	14	9	10	6	4	2	3	4	4	7	9	14	7	14	6	8	5
State average point & area	8	5	6	9	8	14	17	26	14	27	17	18	13	6	4	5	8	8	14	16	25	13	26	14	14	12
County average point & area	21	19	22	33	34	56	77	120	58	63	63	42	65	17	16	19	30	32	55	74	110	55	61	58	35	62
point	31	33	38	37	41	83	35	75	42	78	32	28	40	25	28	33	32	36	80	31	73	35	75	25	20	35
area	21	20	22	33	36	58	82	120	59	63	98	110	77	17	16	20	31	34	56	79	120	56	61	94	100	74
Grid average point & area	32	21	23	33	33	53	77	110	49	52	34	29	38	26	18	21	30	31	52	74	110	46	50	27	22	34
area	23	22	25	36	39	61	86	130	63	65	100	110	78	20	20	22	33	37	60	82	120	59	63	98	110	75
SCC average point	50	55	62	55	61	130	56	120	94	110	110	100	95	41	47	55	47	53	120	49	120	89	100	100	100	90

* With 90% confidence

1 Speciated NOx

2 Speciated VOC

3 Speciated TSP

4 EPA Region

Up to two significant digits have been reported to facilitate comparisons, and does not imply a corresponding level of precision.

- Detailed information on test and other data used to develop emission factors used in Estimation Method 3;
- Comprehensive test data on the variability of fuel sulfur and ash content;
- A detailed statistical analysis of the data available to estimate throughput/activity levels at various geographic levels;
- Sampling data on the use and efficiency of emission control equipment, and a statistical analysis of those data;
- Information on the variation in seasonal, daily, and hourly facility operations. Data on industrial/commercial variations in operations at hourly time periods would be required if the uncertainty in temporal allocation is to be understood;
- Test data on the variation in emissions of various species of pollutants from each SCC;
- Identification of the extent of incomplete data; and,
- The extent of data coding or transcription errors.

While collecting these data would be a major effort, the continued lack of this information will ensure that it is necessary to rely heavily on professional judgment in estimating uncertainty for each of these factors for each source category. This reliance on professional judgment, while not in itself necessarily problematic, reduces the confidence that either analysts or decision-makers can have in the results generated with any method, regardless of its sophistication. It is recognized, however, that to perform the exhaustive research effort required to collect all the input data that are required for the calculation of emissions inventory uncertainty estimates will require more time and money than probably is available for this effort. What is necessary, then, is to collect enough data to provide input values in which users and evaluators of the uncertainty estimates can have confidence. Statistical sampling theory could be applied to ascertain the quantity of data needed to characterize the population with sufficient accuracy.

STATISTICAL METHODS

In addition to the difficulties with data, there are a number of questions concerning the statistical theory used in this study. These questions are focused on the potential influence of bias, covariance and the nature of the observations used in the study. Each of these questions is discussed briefly below.

Bias

It has been assumed in this study that all estimates are unbiased. A systematic bias in an estimate of an emissions parameter, if known, would have the effect of increasing estimates of uncertainty at all levels of aggregation. The effect at the National level could be significant, since there is not the same cancellation of errors that occurs when the random errors have zero mean. One cannot effectively measure a systematic bias in a parameter; if one could, the parameter would be adjusted accordingly to improve the emissions estimates and thereby returning one to the assumption of no bias. A sensitivity analysis should be conducted to investigate the effects of possible unknown systematic biases in the estimated emission factors.

Covariance

The statistical methodology applied in this study does not allow for the possibility that the individual factors included in the uncertainty calculations may not be independent and may, in fact, covary. While it is not certain that these factors do, in fact, covary, it is necessary to at least acknowledge that the possibility exists. Indeed, it seems reasonably likely that factors such as fuel sulfur content and the operating efficiency of control equipment may not be entirely independent of each other. The inclusion of covariance will affect the uncertainty analysis in two ways. First, the formula for the error variance, Equation (3), will include additional terms; and second, the estimate of emission rate given by Equation (2) will be biased, and this bias will need to be carried through the error propagation analysis. Work is required to determine whether and how much the various factors covary, and to assess the impact of any covariance terms on uncertainty estimates.

Nature of Observations

A fundamental assumption in this test of uncertainty methodology is that the emission factors used in the EPA Estimation Method 3, and obtained from AP-42, are the means of a series of representative tests of emissions from a specific source category. However, in actual practice, the emission factors contained in AP-42 often are obtained through the application of professional judgment. Even if the emission factors contained in AP-42 were the mean of a series of tests, it is not evident that the tests included in the AP-42 work are representative of the entire source category. Thus, the assumption that the emission factors are representative of the source category may be fundamentally flawed and the estimates of these emission factors may be biased. Some effort should be expended to either determine the possible impact of the nature of the observations on the statistics or to develop an alternative method of determining uncertainty.

The calculation of the percent error by Equation (1) assumes that the observations are normally distributed. In some cases, the observations are not normally distributed, but are highly skewed and heavy-tailed. The degree

to which the non-normality of the data affects the interpretation of p as calculated by Equation (1) needs to be quantified, and perhaps p should be calculated differently in order to reflect the definition of percent error more accurately.

CONCLUSION

In summary, this study illustrates the results which would be achieved by applying a hypothetical data set to error propagation methodology in estimating the uncertainty of an emissions inventory. The study revealed that the underlying assumptions, methodology, and data need further refinement.

REFERENCES

1. Benkovitz, C. Framework for Uncertainty Analysis of the NAPAP Emissions Inventory. EPA-600/7-85-036 (NTIS PB86-112570), September 1985.

2. Mood, A., F. Graybill, D. Boes. Introduction to the Theory of Statistics, 3rd Edition. McGraw-Hill, 1974.

3. Compilation of Air Pollutant Emission Factors, Fourth Edition, AP-42, (Government Printing Office Stock Numbers 055-000-00251-7, Volume 1 and 055-000-00252-5, Volume 2), U.S. Environmental Protection Agency, Research Triangle Park, NC, November 1985.

4. Preliminary Concepts and Issues Concerning Development of a Data Base of Uncertainty Estimates for Man-Made Sources Task Group. MITRE Working Paper WP-84W485, December 1984.

APPENDIX A

1980 NAPAP EMISSIONS INVENTORY UNCERTAINTY WORKSHOP

Calculation of the uncertainties associated with the emissions values as presented in the NAPAP Emissions Inventory requires estimates of the uncertainties of the individual parameters in the emissions calculations as described in this report. To derive first approximations to the uncertainty values of the parameters used in the calculation of the NAPAP 1980 base year emissions, a 1980 NAPAP Emissions Inventory Uncertainty Workshop was held at the Environmental Protection Agency (EPA) Air and Energy Engineering Research Laboratory (AEERL), Research Triangle Park, North Carolina, on May 7-8, 1985. Since data were unavailable to input into the methodology, engineering judgment was used as the basis to derive a data set to exercise the methodology. An approach was selected for the determination of the uncertainty values based on a concensus of the emissions inventory developers. A panel was assembled for the Workshop; the panel included representation of expertise in:

- emission factor development,
- emission testing,
- development of the 1980 NAPAP Emissions Inventory,
- development of the National Emissions Data System,
- formulation of NAPAP species, spatial, and temporal allocation factors, and
- statistical application of uncertainty estimates to emissions inventories.

The Workshop panelists included:

Carmen Benkovitz, Brookhaven National Laboratory

John Bosch, EPA/OAQPS

Jim Homolya, Radian Corporation

Mike Mills, GCA Corporation

David Mobley, EPA/AEERL

Fred Sellars, GCA Corporation

Ed Trexler, Department of Energy

In addition to the panelists, Dr. Ross Leadbetter, a consultant to the EPA/AEERL in statistics, attended the Workshop as an observer.

Methodology

To accomplish the Workshop objectives, a concensus reaching technique was employed. For each uncertainty value desired, the Workshop participants were asked to assign a value "p" in the following statement:

"90% of the data fall within p% of the mean"

Each participant was then given a prepared questionnaire which addressed specific uncertainty assessment areas. Each question was answered by selecting a value for "p" from a given range of levels for each element of the questions. The uncertainty values to be estimated should not include errors associated with data collection, transcription, and entry. Available values of the parameters should be regarded as "true mean" values; i.e., no bias is introduced by their use. The problems to be studied were outlined as follows:

A. Estimate the uncertainty for SO_2, NO_x, and VOC emission factors of point sources for each estimation method within the 1980 NAPAP Emissions Inventory data base.

   a. Assume the uncertainties for emission factors derived from stack test data are equal for all SCCs.

   b. Assume the uncertainties for emission factors derived from material balances are equal for all SCCs.

   c. Assume that for "Estimation Method 3" existing AP-42 emission factors shall be used with assumed uncertainties assigned to Grades A, B, C, D, and E.

   d. Assume that State and E. H. Pechan derived emission factor uncertainties shall be equal to AP-42 emission factor uncertainties for common SCCs.

   e. Assume that emission factor uncertainties for Estimation Methods 4, 6, 7, 0, and "blank" shall be equal.

   f. Assume 90 percent of the values for an individual source lie within the mean uncertainty estimates.

B. Estimate uncertainties of fuel sulfur content.

   Assume fuel sulfur uncertainties are equal for all SCCs.

C. Estimate uncertainties of production throughput for sources emitting SO_2, NO_x, and/or VOC.

Assume production uncertainties for sources emitting SO_2, NO_x, and VOC shall be equal for all SCCs emitting each pollutant.

D. Estimate uncertainties of control equipment removal efficiencies for SO_2, NO_x, and VOC.

Assume uncertainties for individual control device efficiencies for SO_2, NO_x, and VOC shall be constant across all SCCs.

E. Estimate uncertainties of seasonal, daily, and hourly emissions for point sources of SO_2, NO_x, and VOC.

Assume uncertainties of seasonal, daily, and hourly emissions of each pollutant are the same for all SCCs.

F. Estimate uncertainties of speciation factors for NO_x and VOC.

Assume speciation factor uncertainties are consistent across all SCCs for both point and area sources.

G. Estimate uncertainty introduced by year of record differences for emission of SO_2, NO_x, and VOC.

The adjusted 1980 NAPAP inventory emissions uncertainty equivalent for non-1980 records will be calculated using an adjusted probable error factor which reflects the difference between the emissions year of record and 1980.

H. Estimate uncertainties for SO_2, NO_x, and VOC area source emission factors and activity levels.

Assumptions

a. Area sources are divided into two categories: mobile and others.

b. For the "others" category, emission factor uncertainties for SO_2, NO_x, and VOC are the same for all area source categories.

c. For mobile sources, NO_x and VOC emissions are estimated using appropriate NEDS methodology.

d. All pollutant emission factor uncertainties are identical for all counties.

e. Activity level uncertainties are the same for SO_2, NO_x, and VOC across both the mobile and other categories.

I. Estimate uncertainties of seasonal, daily, and hourly emissions of SO_2, NO_x, and VOC.

Assume pollutant uncertainties are constant across all SCCs.

J. Estimate uncertainties of spatial resolution for area sources of SO_2, NO_x, and VOC.

Assume uncertainties in emissions factors are equal across all counties.

Assume uncertainties in spatial allocation factors are equal across all counties.

After the panel members completed the first iteration of the questionnaire, the responses were assembled and redistributed as a summary to the group. Discussion was limited to clarifying assumptions in the estimating procedure. It should be noted that the uncertainty ranges varied in the initial iteration and did not contain an upper bound. However, the concensus of the panel was to standardize the range for convenience. Thus, values of 10, 25, 50, 75, and 100 percent were selected for the second round.

The questionnaire was then redistributed for a second iteration. An analysis of the summarized responses clearly showed a consensus of the panel members to several elements of the questionnaire. These areas of agreement were identified to the panel and removed from a third and final interation of the questionnaire. All elements of the questionnaire were discussed within the group by noting areas of agreement and disagreement, with the appropriate panel member offering a rationale for his/her response to a particular question. An effort was made by all parties to prevent one individual's opinion from dominating the group's concensus. A consensus was reached on the remaining elements.

Results

The final consensus responses of the expert panel are given in Table A-1; this table follows the questionnaire format.

TABLE A-1.

A. Estimation of uncertainties for SO_2, NO_x, and VOC emission factors of point sources for each estimation method.

Pollutant	Estimation Method	Uncertainty Range (10%) A	(25%) B	(50%) C	(75%) D	(100%) E
SO_2	1. Stack Test		X			
	2. Mat. Bal.	X				
	3. AP-42 Emission Factor Rating					
	A	X				
	B		X			
	C			X		
	D				X	
	E					X
	5. State		X			
	4,6,7,0 and blank					X
NO_x	1. Stack Test		X			
	2. Mat. Bal.		X			
	3. AP-42 Emission Factor Rating					
	A	X				
	B		X			
	C			X		
	D				X	
	E					X
	5. State			X		
	4,6,7,0, and Blank					X
VOC	1. Stack Test		X			
	2. Mat. Bal.			X		
	3. AP-42 Emission Factor Rating					
	A	X				
	B		X			
	C			X		
	D				X	
	E					X
	5. State			X		
	4,6,7,0, and blank					X

(continued)

TABLE A-1 (continued)

B. Estimation of uncertainties for fuel sulfur content.

	Uncertainty Range				
	(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
Fuel sulfur content	X				

C. Estimation of uncertainties of production throughput for sources emitting SO_2, NO_x, and/or VOC.

	Uncertainty Range				
	(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
SO_2		X			
NO_x		X			
VOC		X			

(15% for SO_2, NO_x, and VOC)

D. Estimation of uncertainties of control equipment removal efficiencies for SO_2, NO_x and VOC.

	Uncertainty Range				
	(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
SO_2 removal efficiency		X			
VOC removal efficiency		X			
NO_x removal efficiency		X			

(continued)

TABLE A-1 (continued)

E. Estimation of uncertainties of seasonal, daily, and hourly emissions for point sources of SO_2, NO_x, and VOC.

		Uncertainty Range				
		(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
Seasonal Emissions	SO_2	X				
	NO_x	X				
	VOC	X				
Daily Emissions	SO_2		X			
	NO_x		X			
	VOC		X			
Hourly Emissions	SO_2			X		
	NO_x			X		
	VOC				X	

F. Estimation of uncertainties of speciation factors for NO/NO_2, and VOC.

	Uncertainty Range				
	(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
NO/NO_2 speciation		X			
VOC speciation					X

(continued)

TABLE A-1 (continued)

G. Estimation of error introduced by year-of-record differences for emissions of SO_2, NO_x, and VOC.

Year of Record	Year of Record Difference Uncertainty (10%) A	(25%) B	(50%) C	(75%) D	(100%) E
1969					X
1970					X
1971				X	
1972				X	
1973			X		
1974			X		
1975		X			
1976		X			
1977		X			
1978	X				
1979	X				
1981	X				
1982	X				
1983		X			
1984		X			

(continued)

TABLE A-1 (continued)

H. Estimation of uncertainties for SO_2, NO_x, and VOC area source emission factors.

		Uncertainty Range				
		(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
Mobile Sources	NO_x			X		
	VOC					X
"Other" Sources	SO_2		X			
	NO_x		X			
	VOC					X
Activity Level	Mobile		X			
	Other		X			

I. Estimation of uncertainties of seasonal, daily, and hourly emissions for area sources of SO_2, NO_x, and VOC.

		Uncertainty Range				
		(10%) A	(25%) B	(50%) C	(75%) D	(100%) E
Seasonal Emissions	SO_2	X				
	NO_x	X				
	VOC	X				
Daily Emissions	SO_2		X			
	NO_x		X			
	VOC		X			
Hourly Emissions	SO_2			X		
	NO_x			X		
	VOC				X	

(continued)